Federated Learning for Industrial Entity Extraction

Shengze Fu, Xiaoli Zhao(✉), and Chi Yang

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China
1012439688@qq.com

Abstract. Entity extraction in the industrial field is an important part of the realization of digital transformation in the industrial field. The construction of entity extraction model in the industrial field requires a large amount of data from various parties. However, due to the security and privacy issues of the data, the data in the industrial field often exists in the form of islands, so it is almost impossible to integrate the data scattered among various parties. Therefore, this paper proposes a federated learning framework to assist parties in industry to overcome data silos and collaborate in building entity extraction models. The solution to the Non-IID problem in federal learning is to find an index to measure the data performance of all participants. Participants with relatively good data performance have a higher weight in the aggregation stage, while participants with relatively poor data performance have a lower weight in the aggregation stage. In this paper, an aggregation update method FedData is proposed to improve the performance of federated learning in data Non-IID scenarios. The method measures the data performance of each participant based on the aggregate test performance of each participant's local model on the private data of other participants and assigns aggregate weights to each participant based on this. The experimental results show that the framework can make the participants who cannot cooperate in modeling jointly build the entity extraction model without being constrained by data security and privacy issues, so as to achieve better results. Moreover, the aggregation update method proposed in this paper has better performance than FedAvg in the scenario where the data is not independent and equally distributed.

Keywords: Industrial field · Entity extraction · Federated learning · Non-IID

1 Introduction

Knowledge graph is a visualization technology to display knowledge architecture and knowledge points in information [1], which is originally intended to improve users' search experience. The basic unit of knowledge graph is the triplet composed of "entity-relation-entity", which is also the core of knowledge graph, and its essence is a huge semantic network graph. In recent years, more and more fields are interested in using knowledge graph technology. As the society attaches more and more importance to knowledge graph technology, people have made a lot of progress in the research of

G. Zhai et al. (Eds.): IFTC 2022, CCIS 1766, pp. 490–501, 2023.
https://doi.org/10.1007/978-981-99-0856-1_36

knowledge graph [2–4]. Entity extraction is the first and key step in the construction of knowledge map, so this paper mainly studies the entity extraction in the industrial field, which will be helpful to the construction of knowledge map in the industrial field.

Entity extraction is also known as named entity recognition [5], whose main task is to identify named entities in text and classify them into predefined entity categories. Entity extraction is an important part of knowledge graph construction. The key problem is how to build a high-quality entity extraction model to extract the desired entity information from massive data sources. Therefore, the research on entity extraction in industrial domain is conducive to the construction of industrial domain knowledge graph.

The industrial field involves a large number of data acquisition and analysis operations such as industrial equipment fault monitoring and pattern sensing. The real time and complexity of industrial equipment data are not qualified by traditional database technology. Therefore, knowledge graph, a technology to show the relationship between data structures, has been applied more and more widely in industry. The construction of knowledge graph in the industrial field requires a large amount of data from all parties. However, as oil in the industrial field, data is an important resource, and the data of all parties often cannot be shared due to the business competition and security and privacy concerns of the industry itself, thus forming data islands [6]. Traditional machine learning methods integrate data for unified machine learning training, but this approach has the risk of data leakage. In order to solve this problem, Google proposed a federated learning solution to jointly model and share computing results under the premise of protecting the privacy and security of the original data [7, 8]. Although federated learning can effectively solve the problem of data islanding, in many practical scenarios, the data of all parties in federated learning are usually non-independent and identically distributed (called Non-IID). Literatures [9, 10] have shown through various experiments that Non-IID data will seriously affect the performance of federated learning.

Aiming at the problem of data islanding in the industrial domain, this paper proposes a federated learning framework for entity extraction in the industrial domain, which makes the parties that cannot cooperate with each other due to data security and privacy issues participate in modeling jointly, thus improving the model. In the experimental part, BERT + BiLSTM + CRF model is used as the entity extraction model, which verifies the feasibility of the federated learning framework applied to entity extraction in the industrial field, and an aggregation update method FedData is proposed to improve the performance of federated learning in Non-IID scenarios. The method adjusts the weight of the local model in the global model according to the comprehensive test accuracy of each participant's local model on the private data of all other participants. This paper uses the industrial equipment failure order data of an automobile group. The main contributions are as follows:

(1) The effectiveness of the federated learning framework is verified on the failure work order data of industrial equipment of an automobile group, and the feasibility of FedData method is verified under two Non-IID scenarios.
(2) Aiming at the Non-IID problem of data in the industrial field, this paper proposes a federated learning aggregation update algorithm FedData to improve the performance of the federated learning framework in the face of non-independent and identically distributed data in the industrial field.

(3) The federated learning is applied in the industrial field, and a federated learning framework is proposed to jointly construct entity extraction models under the premise of ensuring data security, which lays a foundation for the construction of knowledge graph in the industrial field.

2 Relevant Content

2.1 Entity Extraction

Entity extraction aims to automatically identify desired entities from unstructured text and label them into predefined categories, such as people, places, and organizations. Common annotation methods are BIO annotation: B-begin, I-inside, O-outside; B represents the beginning of the entity, I represents the content of the entity, and O represents the non-entity part. Since entity is the most basic element in knowledge graph, the completeness and accuracy of its extraction will directly affect the quality of knowledge base. Therefore, entity extraction is the most basic and key step to construct knowledge graph.

In the early days, entity extraction methods were based on statistical learning and rules. Although the traditional methods have achieved good results, they are highly dependent on professional domain knowledge and difficult to construct artificial features. Aiming at the problems existing in traditional methods, Hammerton et al. [11] first applied neural networks to the research of named entities. They used one-way long short-term memory network (LSTM), which has good sequence modeling ability, so LSTM-CRF is considered as the infrastructure of entity extraction. Later, on the basis of this architecture, Guillaume Lample et al. [12] proposed a neural network model combining Bidirectional Long Short-Term Memory (BiLSTM) and conditional random field (CRF). This architecture can extract the sequence information of the context, so it is widely used in the field of entity extraction. However, this method pays attention to the feature extraction of words, characters or between words, but ignores the context or semantics of words, which leads to the poor effect of entity extraction. In order to solve the above problems, Jacob Devlin et al. [16] used BERT (Bidirectional Encoder Representation from Transformers) language preprocessing model to represent word vectors. BERT can fully describe the relationship characteristics between characters, words and even sentences, and better represent the contextual and semantic information in different contexts.

In recent years, deep learning-based methods have been widely applied to Chinese named entity recognition research [13–16]. Compared with traditional methods, deep learning-based methods can learn independently from original data and find deeper and more abstract features, which has the advantage of stronger generalization.

2.2 Federated Learning

With the rapid development of digital society, technologies related to artificial intelligence and big data have been attached great importance, which not only bring new development opportunities for traditional industries, but also inevitably bring data security and privacy problems, and data island problem is one of the key problems. Federated

Learning provides a solution to the current problems faced by artificial intelligence and big data industry. This technology can complete joint modeling while protecting the data privacy of all parties.

Federated learning is a basic technology of artificial intelligence, which is essentially a machine learning framework. Its original intention is to assist multiple participants or multiple computing nodes to carry out machine learning to achieve the purpose of security modeling on the premise of ensuring data of all parties, protecting privacy and security, and ensuring legitimacy. Federated learning adheres to the idea of "only passing model parameters or gradients". Data of all parties only needs to be kept locally, thus avoiding data leakage.

As a user-based distributed machine learning method, federated learning has many advantages. It can directly train effective machine learning models based on users' local data, and make full use of high-quality data from various parties. The federated learning framework mainly includes data holders and central servers. The federated learning framework is mainly divided into the following steps, as shown in Fig. 1:

(1) Initialization: All users get an initialization model from the central server, they can join the federated learning, and determine the same task and model training objectives.
(2) Local computation: In the communication process of each round of federated learning, federated learning users first get the global model parameters from the central server, and then use their private training samples to train the model, update the model, and send these updates to the central server.
(3) Central aggregation: the global model of the next round can be obtained by aggregating the models trained by different users and updating them.
(4) Model update: The central server updates the global model once according to the aggregated results, and returns the updated model to the data holders participating in federated learning.

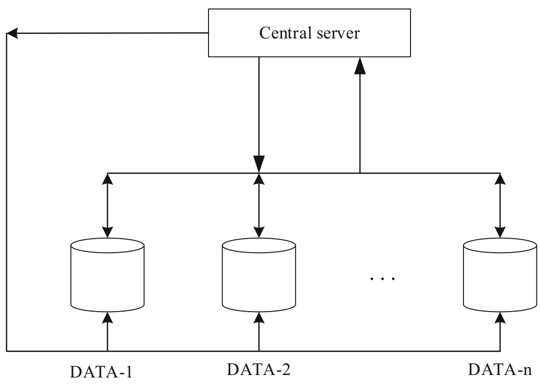

Fig. 1. Federal learning framework

Federated Average Algorithm (FedAvg) [17] is the most common algorithm scheme in the federated learning framework. The improvement and theoretical analysis of FedAvg algorithm is an important research direction in current federated learning [18–21]. Federated learning still faces some problems and challenges [22].

3 Federated Learning Framework for Industrial Entity Extraction

3.1 Entity Extraction Model

At present, methods based on deep learning can achieve better results, so this study uses BERT + BiLSTM + CRF as the entity extraction model, and uses the industrial equipment failure order data of an automobile group as corpus data. The model is divided into three modules: BERT (Bidirectional Encoder Representation from Transformers), namely Encoder of bidirectional Transformer; BiLSTM is composed of forward LSTM and backward LSTM. CRF is a conditional random field. The model structure is shown in Fig. 2.

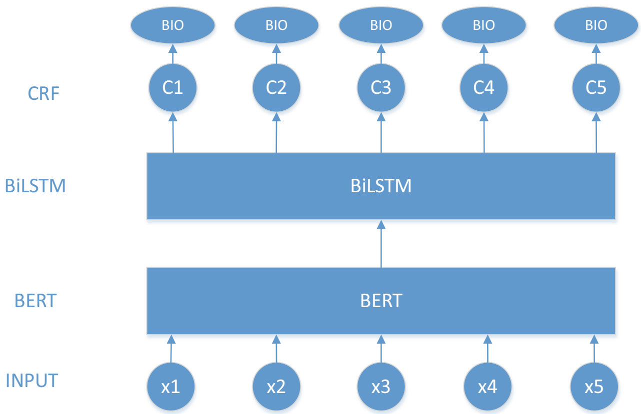

Fig. 2. Structure diagram of BERT + BiLSTM + CRF model

The workflow of the whole model is as follows: Input the corpus data of the industrial equipment fault order of the automobile group as the training data. Firstly, the BERT pre-training model is used to obtain the word vector and extract the important features of the text. Then, BiILSTM deep learning context feature information is used for named entity recognition. Finally, CRF layer processes the output sequence of BiLSTM to get a predicted annotated sequence, and then extracts and classifies each entity in the sequence.

3.2 FedData

In each round of FedAvg, the server first sends the global model to each participant, then each party updates the model with its local data set, and then sends the updated model back to the server. Finally, the server receives the local model and performs aggregation update, and finally gets the next round of global model. In the model aggregation stage of FedAvg, the weight of each participant is determined according to the proportion of its own data volume in the total data volume. However, the contribution of each participant to the global model is not necessarily positively correlated with the data volume, but also affected by data distribution and data quality. Since the principle of "only model parameters or gradients are passed" is always held in the process of federated learning, the method in this paper takes advantage of this feature to propose the FedData method, which can only pass the local model between different participants without data leakage. The starting point was to find a better metric to guide weight allocation in the model aggregation phase instead of data volume. Before the federal learning framework is fully developed, each participant uses the initial model for training and transmits the local training model to each other. The comprehensive test performance of each participant's local model on the private data of other participants is used to measure the data performance of each participant. Participants with relatively good data performance have a higher weight in the aggregation stage, while participants with relatively poor data performance have a lower weight in the aggregation stage. The FedData method is described as follows:

There are n participants in total. First, participant i trains the local model w_i, which is transmitted to all other participants through the central server, and the model testing accuracy is $v_{i,j}$ on the data of participant j. Then the data index of participant i can be expressed as the average value of all $v_{i,j}$ values, as shown in Eq. (1):

$$T_i = \frac{\left(v_{i,1} + v_{i,2} + \ldots + v_{i,n-1}\right)}{n-1} \tag{1}$$

$D(i)$ represents the weight of participant i in the federated learning model aggregation stage. In FedData, the weight of each participant is shown in formula (2):

$$D(i) = \frac{T_i}{(T_1 + T_2 + \ldots + T_n)} \tag{2}$$

The overall federated learning algorithm is shown in Algorithm 1.

Algorithm 1：FedData

Input: number of participants n, number of global model update rounds T, number of local iteration rounds M, Global model W, local model w, learning rate η

Output: the final model W_{T}

Server:
initialize the global model W_0
for t= 1, 2, …, T **do**
global model W_{t} is sent to each participant
for participant $i \in n$ **do**
initial model of the participant $w_t^i = W_{\mathrm{t}}$
$w_{t+1}^i \leftarrow w_t^i$ local update
end
$W_{\mathrm{t+1}} \leftarrow \sum_{i=1}^{i=n} D(i)\, w_{t+1}^i$
end

Client:
w_t^i local update:
for m=1,2, …, M **do**
$w_{m+1}^i \leftarrow w_m^i - \eta \triangledown f(w_m^i)$
return $W_{\mathrm{t+1}}$ to the server to get a new round of global model

The federated learning framework for entity extraction in the industrial domain is shown in Fig. 3, which shows the process of the framework and the help of using the framework to build the knowledge graph in the industrial domain. The work of the framework is mainly divided into the following steps.

(1) Initialization: First, the central government and all participants determine the common training objectives, and the central server initializes the original model. This framework adopts the BERT + BiLSTM + CRF joint model.
(2) Distribution model: the central server sends its own model to each participant, and each participant receives the model from the central server.
(3) Local training: Participants in federated learning get the global model from the central server, then use their private data to train the model locally and update the model parameters, and send these updates to the central server.
(4) Model aggregation: the central server uses FedData to aggregate and update the model parameters sent by each participant to obtain the global model of the next round; The above steps 2, 3 and 4 were repeated to optimize the global model. After the training, the joint model trained by all parties was obtained.

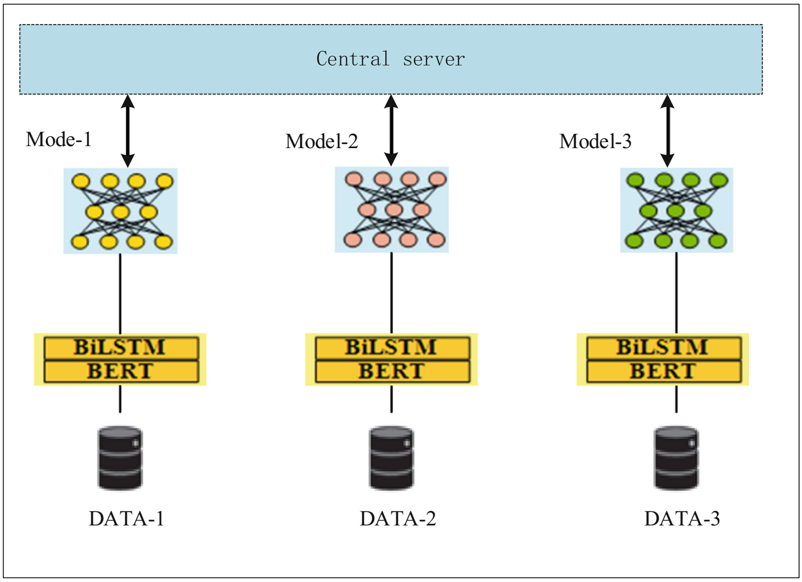

Fig. 3. Federated learning framework for industrial entity extraction

4 Experiment

4.1 Introduction of Data

The data in the experiment is the work order data of industrial equipment failure of an automobile group. Firstly, entities are annotated. There are six entity labels: ATTRIBUTE, NORMAL, UNNORMAL, FAULT, DEVICE, OPERATION, and non-entity label 0. The sample is shown in Fig. 4. In traditional machine learning, the data is distributed on the same machine, and it is assumed that the data are sampled independently from the same distribution, that is, the data in traditional machine learning is independent and identically distributed. However, in the federated learning scenario in the industrial field, because the equipment belongs to a certain enterprise, factory, or department, the data distribution is often very different, that is, the data is not independent and identically distributed. The amount of data owned by different data holders and the label category of data are very different, so the experiment in this paper will process the data from the perspectives of data amount and label category to simulate the scenario of Non-IID.

In this paper, two different processing methods are used to select data from the training set to construct two kinds of Non-IID scenarios, indicating the different degree of heterogeneity of data, which are named as low Non-IID and high Non-IID respectively. Among them, the selection method of low Non-IID is to randomly intercept multiple blocks of data from the original data set and distribute them to different participants, and the total amount of data of each participant is significantly different. The high Non-IID selects the data of partial labels from the original data and assigns them to each

机	B-DEVICE
械	I-DEVICE
手	I-DEVICE
无	B-UNNORMAL
法	I-UNNORMAL
上	I-UNNORMAL
电	I-UNNORMAL
，	O
检	B-OPERATION
查	I-OPERATION
后	O

Fig. 4. Data sample

participant. There are obvious differences in the total amount of data contained by each participant and the types of labels.

For the low Non-IID scenario, three sub-datasets low1, low2 and low3 are set to represent the private data of the three parties, among which low1 has 1000 rows of data, low2 has 875 rows of data, and low3 has 525 rows of data. For the high Non-IID scenario, three sub-datasets, high1, high2 and high3, are set up to represent the private data of the three participants. high1 has 100 lines of data, and only the data of O, DEVICE, and UNNORMAL labels are available. There are 125 lines of data in high2, only the data of O, DEVICE, NORMAL and OPERATION labels; There are 150 lines of data in high3, and only five labels of O, DEVICE, FAULT, OPERATION, and ATTRIBUTE are available.

4.2 Experimental Result

In order to verify the reliability of the federated learning framework and the effect of FedData, the experiment compares the effects of the participant training alone and each participant using the federated learning framework, and the effects of using FedAvg and FedData respectively in two scenarios: low Non-IID and high Non-IID.

Based on the experiment and formula (1), in low Non-IID and high Non-IID scenarios, the data indexes T_1, T_2, T_3 are shown in Table 1. These indexes are used to represent the data quality of each participant.

Table 1. Data index

Compare objects	T_1	T_2	T_3
Low Non-IID	0.65	0.60	0.63
High Non-IID	0.24	0.51	0.50

As can be seen from the data indicators, the data indicators of all participants in the low Non-IID scenario are close, and the weight of each participant is similar in the aggregation stage. In the high Non-IID scenario, for example, if the data index of Participant 1 is low, the weight of Participant 1 will be reduced in the aggregation phase. According to Formula (2), the aggregated weight values $D(1)$, $D(2)$ and $D(3)$ of each participant using FedAvg and FedData in the two data scenarios are shown in Table 2.

Table 2. Weight parameter value

Compare objects	$D(1)$	$D(2)$	$D(3)$
FedAvg-low Non-IID	$\frac{1000}{2400}$	$\frac{875}{2400}$	$\frac{525}{2400}$
FedData-low Non-IID	$\frac{65}{188}$	$\frac{60}{188}$	$\frac{63}{188}$
FedAvg-high Non-IID	$\frac{100}{375}$	$\frac{125}{375}$	$\frac{150}{375}$
FedData-high Non-IID	$\frac{24}{125}$	$\frac{51}{125}$	$\frac{50}{125}$

As can be seen from Table 2, the model aggregation weight of FedData is quite different from that of FedAvg. The amount of data of participants cannot well reflect the data quality of each participant, and the method allows participants with relatively good data performance to have higher weight in the aggregation stage. Table 3 and Table 4 record the experimental results under low Non-IID and high Non-IID scenarios respectively.

Table 3. Experimental evaluation of low Non-IID scenarios

Compare objects	Accuracy
low1	64.9
low2	69.6
low3	66.7
FedAvg	72.7
FedData	73.2

It can be seen from the experimental results that after using the federated learning framework, the effect of joint modeling of all parties is significantly better than that

Table 4. Experimental evaluation of high Non-IID scenarios

Compare objects	Accuracy
high1	37.9
high2	53.4
high3	67.1
FedAvg	69.4
FedData	70.2

of individual modeling of all parties in both low Non-IID and high Non-IID scenarios. In addition, the accuracy of FedData in low Non-IID and high Non-IID scenarios is improved by 0.5% and 0.8% compared with FedAvg, respectively. Experiments show that this method has better performance than FedAvg in Non-IID scenarios, and can alleviate the impact of Non-IID on federated learning.

5 Conclusion

In order to help all parties in the industrial domain to jointly model under the premise of protecting data privacy, this paper proposes a federated learning framework for entity extraction in the industrial domain, and puts forward FedData method to improve the performance of federated learning on Non-IID data. Experiments on different datasets show that the proposed framework and method are effective in industrial entity extraction. Although federated learning can be widely applied in the industrial field, it still faces many problems and challenges, which can be summarized as the following aspects: first, the heterogeneity of federated learning; Second, communication efficiency; Third, privacy and security. In the future, federal learning research in the industrial field needs to further explore these issues.

References

1. Singhal, A., et al.: Introducing the knowledge graph: things, not strings. Official Google Blog **5**, 16 (2012)
2. Cai, H., Zheng, V.W., Chang, K.C.-C.: A comprehensive survey of graph embedding: problems, techniques, and applications. IEEE Trans. Knowl. Data Eng. **30**(9), 1616–1637 (2018)
3. Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q.: Ernie: enhanced language representation with informative entities. arXiv preprint arXiv:1905.07129 (2019)
4. Liu, W., et al.: K-BERT: enabling language representation with knowledge graph. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 2901–2908 (2020)
5. Zhao, S., Cai, Z., Chen, H., Wang, Y., Liu, F., Liu, A.: Adversarial training based lattice LSTM for Chinese clinical named entity recognition. J. Biomed. Inform. **99**, 103290 (2019)
6. Jolfaei, A., Ostovari, P., Alazab, M., Gondal, I., Kant, K.: Guest editorial special issue on privacy and security in distributed edge computing and evolving IoT. IEEE Internet Things J. **7**(4), 2496–2500 (2020)

7. Kairouz, P., et al.: Advances and open problems in federated learning. Found. Trends Mach. Learn. **14**(1–2), 1–210 (2021)
8. Chen, Y., Sun, X., Jin, Y.: Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation. IEEE Trans. Neural Netw. Learn. Syst. **31**(10), 4229–4238 (2019)
9. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., Chandra, V.: Federated learning with non-IID data. arXiv preprint arXiv:1806.00582 (2018)
10. Sattler, F., Wiedemann, S., Müller, K.-R., Samek, W.: Robust and communication-efficient federated learning from non-IID data. IEEE Trans. Neural Netw. Learn. Syst. **31**(9), 3400–3413 (2019)
11. Hammerton, J.: Named entity recognition with long short-term memory. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, pp. 172–175 (2003)
12. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360 (2016)
13. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
14. Li, J., Sun, A., Han, J., Li, C.: A survey on deep learning for named entity recognition. IEEE Trans. Knowl. Data Eng. **34**(1), 50–70 (2020)
15. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
16. Dong, C., Zhang, J., Zong, C., Hattori, M., Di, H.: Character-based LSTM-CRF with radical-level features for Chinese named entity recognition. In: Lin, C.-Y., Xue, N., Zhao, D., Huang, X., Feng, Y. (eds.) ICCPOL/NLPCC -2016. LNCS (LNAI), vol. 10102, pp. 239–250. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50496-4_20
17. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282. PMLR (2017)
18. Wu, W., He, L., Lin, W., Mao, R., Maple, C., Jarvis, S.: SAFA: a semi-asynchronous protocol for fast federated learning with low overhead. IEEE Trans. Comput. **70**(5), 655–668 (2020)
19. Lu, X., Liao, Y., Lio, P., Hui, P.: Privacy-preserving asynchronous federated learning mechanism for edge network computing. IEEE Access **8**, 48970–48981 (2020)
20. Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., Khazaeni, Y.: Federated learning with matched averaging. arXiv preprint arXiv:2002.06440 (2020)
21. Chai, Z., Chen, Y., Anwar, A., Zhao, L., Cheng, Y., Rangwala, H.: FedAT: a high-performance and communication-efficient federated learning system with asynchronous tiers. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1–16 (2021)
22. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: challenges, methods, and future directions. IEEE Signal Process. Mag. **37**(3), 50–60 (2020)